DOCUMENT RESUME

ED 229 942 EC 151 951

AUTHOR Vinsonhaler, John F.; And Others
TITLE Diagnosing Children with Educational Problems: Characteristics of Reading and Learning Disabilities Specialists and Classroom Teachers.
INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
REPORT NO IRT-RS-117
PUB DATE Oct 82
CONTRACT 400-81-0014
NOTE 58p.
AVAILABLE FROM Michigan State University, Institute for Research on Teaching, College of Education, 252 Erickson Hall, East Lansing, MI 48824 ($5.00).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Clinical Diagnosis; Elementary Secondary Education; *Handicap Identification; *Learning Disabilities; Reading Consultants; *Reading Difficulties; Reading Teachers; Simulation; Special Education Teachers; Student Evaluation; Teachers

ABSTRACT

Six studies in laboratory and classroom settings investigated the diagnostic and remedial performance of 66 reading and learning disabled specialists and classroom teachers. Ss were asked to diagnose simulated cases of either reading or learning disability and suggest an initial remediation plan. Two related findings were revealed across all studies. First, commonality (the extent to which clinicians made the same statements about a case) was very low; most statements in the written diagnoses and remediations for a given case were mentioned only once. Only 3% of the statements were mentioned in half or more of the diagnoses for the same case. Second, individual agreement (between two clinicians on the same case and one clinician at two different times for the same case) was also very low. Mean diagnostic agreement between two clinicians remained close to zero across all studies. Mean diagnostic agreement results for a single clinician on a case across time showed that only 20 percent of the statements were agreed upon both times by the same person. Additionally, analysis of diagnostic and remedial process in three of the studies revealed wide variability in total time taken to collect case information (cues) and in number of cues collected. Neither were significantly correlated with agreement. (Author/CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

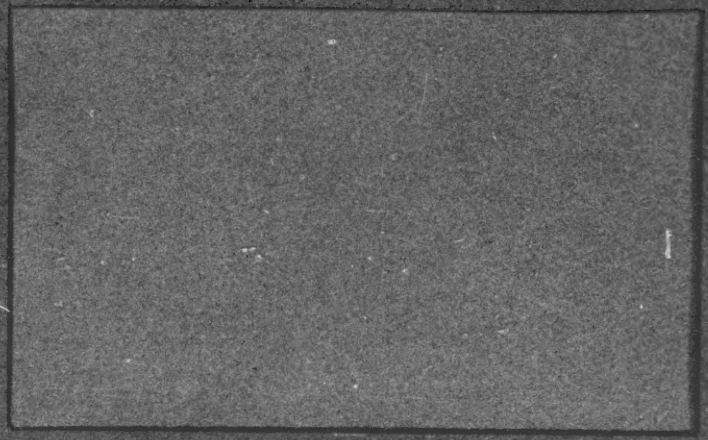
ED229942

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.



Institute for Research on Teaching

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John F. Vinsonhaler

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Research Series No. 117

DIAGNOSING CHILDREN WITH EDUCATIONAL PROBLEMS: CHARACTERISTICS OF READING AND LEARNING DISABILITIES SPECIALISTS AND CLASSROOM TEACHERS

John F. Vinsonhaler, Annette B. Weinshank,
Christian C. Wagner, and Ruth M. Polin

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

October 1982

Publication of this work is sponsored by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Program for Teaching and Instruction of the National Institute of Education, United States Department of Education. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-81-0014)

Institute for Research on Teaching

The Institute for Research on Teaching was founded at Michigan State University in 1976 by the National Institute of Education. Following a nationwide competition in 1981, the NIE awarded a second contract to the IRT, extending work through 1984. Funding is also received from other agencies and foundations for individual research projects.

The IRT conducts major research projects aimed at improving classroom teaching, including studies of classroom management strategies, student socialization, the diagnosis and remediation of reading difficulties, and teacher education. IRT researchers are also examining the teaching of specific school subjects such as reading, writing, general mathematics, and science, and are seeking to understand how factors outside the classroom affect teacher decision making.

Researchers from such diverse disciplines as educational psychology, anthropology, sociology, and philosophy cooperate in conducting IRT research. They join forces with public school teachers, who work at the IRT as half-time collaborators in research, helping to design and plan studies, collect data, analyze and interpret results, and disseminate findings.

The IRT publishes research reports, occasional papers, conference proceedings, a newsletter for practitioners, and lists and catalogs of IRT publications. For more information, to receive a list or catalog, and/or to be placed on the IRT mailing list to receive the newsletter, please write to the IRT Editor, Institute for Research on Teaching, 252 Erickson Hall, Michigan State University, East Lansing, Michigan 48824-1034.

Co-Directors: Jere E. Brophy and Andrew C. Porter

Associate Directors: Judith E. Lanier and Richard S. Prawat

Editorial Staff
Editor: Janet Eaton
Assistant Editor: Patricia Nischan

Abstract

A series of six studies conducted in laboratory and classroom settings investigated the diagnostic and remedial performance of reading and learning disabilities specialists and classroom teachers. The participants' (N=66) basic task was to diagnose simulated cases of either reading or learning disability, and to suggest an initial remediation plan. There are two related findings across all studies. First, commonality (the extent to which clinicians made the same statements about a case) is very low; most statements in the written diagnoses and remediations for a given case were mentioned only once. Only 3% of the statements were mentioned in half or more of the diagnoses for the same case. Second, individual agreement (between two clinicians on the same case and one clinician at two different times for the same case) was also very low. Mean diagnostic agreement between two clinicians remained close to zero across the six studies. Mean diagnostic agreement results for a single clinician on a case across time showed that only 20% of the statements were agreed upon both times by the same person. Additionally, analysis of diagnostic and remedial process in three of the studies revealed wide variability in total time taken to collect case information (cues) and in the number of cues collected. Neither were significantly correlated with agreement.

DIAGNOSING CHILDREN WITH EDUCATIONAL PROBLEMS: CHARACTERISTICS OF READING AND LEARNING DISABILITIES SPECIALISTS AND CLASSROOM TEACHERS

John F. Vinsonhaler, Annette B. Weinshank, Christian C. Wagner, and Ruth M. Polin[1]

Diagnosis is accorded importance by nearly all authorities in the field of reading. Diagnosis as the basis for remediation is a cardinal principle in the literature and in the world of practice (Ekwall, 1976; Spache & Spache, 1973; Carter & McGinnis, 1970; Otto, McMenemy, & Smith, 1973; Rabinovitch, 1965; Smith, 1969; Smith, Carter, & Dapper, 1970). Many view diagnosis as an essential and integral part of total reading instruction and as a basic element of all efficient teaching (Otto et al., 1973; Sheldon, 1968; Smith et al., 1970). Diagnosis is seen as a preliminary step to sound instruction; a guide to teachers in the planning, modification, and individualization of instruction (Bond, 1970; Bond & Tinker, 1967; Dauzat, 1977; Dietrich, 1972; Farr, 1971; Karlsen, 1976; Olson & Dillner, 1976; Sawyer, 1968; Smith, 1969; Smith et al., 1970; and Swalm, 1973; Austin, Note 1).

While it is generally agreed that diagnosis is important, there is less consensus on its content, how it is conducted, and the frequency of a useful diagnosis.

Major Orientations

At least three major orientations toward diagnostic content can be found in the literature. One approach concentrates on establishing the

---

[1]John Vinsonhaler is co-coordinator with Annette Weinshank of the Outcomes in Reading Project at the IRT. Vinsonhaler is a professor in the Counseling, Educational Psychology, and Special Education Department; Annette Weinshank is a teacher collaborator with the IRT. Christian Wagner is a consultant to the project. Ruth Polin is data-processing coordinator for the Clinical Studies Project.

child's general reading level as compared to his/her reading potential (Guszak, 1972; Spache, 1976). A second orientation emphasizes the examination of the child's performance on a set of reading skills. Some authors suggest that the diagnosis include both strengths and weaknesses (Peters, 1977; Monroe, 1968; Carter, 1970; Carter & McGinnis, 1970). A third group of authors view the diagnosis as a determination of causality, that is, an understanding of the underlying factors that have caused the reading problems. Such an understanding, they feel, enables the clinician to prescribe the most appropriate steps for remediation (Harris, 1972; Strang, 1964; Natchez, 1968; Monroe, 1968; Carter & McGinnis, 1970; Harris, 1977).

More specifically, the first method emphasizes a diagnosis conducted by the teacher in the classroom and concerns the early detection of reading problems. This type involves little clinical testing or interaction with individual students. Classroom diagnosis is typically a group event involving the administration of group tests (Carter & McGinnis, 1970; Kennedy, 1971; Otto et al., 1973; Smith et al., 1970; Wilson, 1977). As such, it does not require much time and is an informal process in which the classroom teacher can observe a group or individual students over a long period of time (Smith et al., 1970; Wilson, 1977).

The second method of diagnosis posits that reading difficulties of some students are too serious to be dealt with solely by the classroom teacher. A specialist becomes responsible for diagnosis (Smith et al., 1970; Wilson, 1977). Such a diagnosis focuses mainly on skill performance and is formal, analytical, and specific (Bond & Tinker, 1967).

A third method emphasizes a diagnosis performed in a reading clinic. Clinical diagnosis of reading difficulties is designed to deal with severe cases that cannot be handled in a regular school setting. Although part of this diagnosis can be conducted by a school reading specialist, other phases must be carried out by clinicians from various disciplines (psychologists, audiologists, physicians, etc.). Clinical diagnoses are oriented mainly toward the determination of causal factors. They require an intensive, thorough case study of an individual child, including personality factors (Strang, 1969).

Frequency of Diagnosis

Some authors argue that the diagnosis should be conducted on a regular basis before and during remediation. Others argue that diagnosis should be a continuous process in response to changing information about the child and his reading problems (Bond, 1970; Bond & Tinker, 1967; Otto, et al., 1973; Spache, 1976; Smith, 1969; Smith, et al., 1970; Strang, 1964). While empirical evidence on optimum frequency of reading diagnosis is scarce, research on physicians' decision making confirms the view that diagnosis occurs over time and is modified in the face of new data. Eventually, however, for most physicians, the diagnosis stabilizes to form the basis of an initial plan of therapy. Regardless of the method, content, and frequency of reading diagnosis, nearly all authors agree that the diagnosis should form the basis for remediation. Here again, however, little empirical evidence exists in reading about the relationship between individual diagnosis and remediation. Spache (in Newman, 1969) contended that there was still widespread lack of integration between the two processes of diagnosis and remediation and stated,

> Numerous reports of remedial work give evidence that the procedures used are not directly related to the detailed diagnostic findings.
>
> Bateman (Note 2) asked,
>
> Was the diagnosis a necessary and sufficient prerequisite to the remediation? Might other remediation, not derived from the diagnosis, have been equally successful? A child is diagnosed but the remediation is not successful. Was the diagnosis inadequate, or was an error made in deriving the remediation?

Instructional Consequences of Low Reliability

Bateman's point can be made more concrete by examining the instructional consequences of unreliable diagnoses for the same or similar cases. Low reliability can be interpreted as a function of unpredictable judgments, that is, the chance assignment of children to diagnostic categories. First, consider the case in which children have a problem with word attack skills and a known effective remediation is available. In the context of low diagnostic reliability, some of the children will be diagnosed as having the problem, will be correctly treated, and will show improvement in reading. The other children in the group will be incorrectly diagnosed, will receive no remediation and will show an overall loss in performance in relation to their classmates.

Now consider the case of a group of children who do not have a word attack problem. Some will be correctly diagnosed as not having word attack problems and will receive no treatment. Others will be incorrectly diagnosed and will spend their academic time on drill and practice for skills already mastered.

In the examples above, the effectiveness of the remediation was known. Now, let us examine the impact of unreliable diagnosis when the efficacy of treatment is not known and must be evaluated. An

evaluation study is performed by obtaining the apparent diagnosis for each child and applying the correct remediation in terms of the stated diagnosis. Suppose a group of children cannot read two- and three-syllable words. Within this group some children lack a mastery of major sound-symbol associations (sound-symbol problem). Others have poor syllabication and blending skills (syllabication problem). Further, suppose we are evaluating two treatments: Treatment A is effective for the sound-symbol problem and Treatment B is effective for the syllabication problem. Assume that each treatment works primarily for one problem only.

Consider the group that receives the apparent diagnosis of sound-symbol problem. Some of the group will actually have this problem, will receive Treatment A, and will show good improvement. The others will actually have the syllabication problem, will receive Treatment A instead of Treatment B, and will show no improvement. Overall, the group with the apparent diagnosis of sound-symbol problem will show only a modest improvement as a result of Treatment A. A similar dilution of treatment effect will obtain for the group with the apparent syllabication problem. Treatment B will not be appropriate for the entire group.

Overall, the effectiveness of the two treatments will be systematically underestimated to the degree that the diagnoses are unreliable. Obviously, reliability of diagnosis provides no information as to its validity (one can be reliably wrong). Reliability only permits the correct estimation of remedial effectiveness (Collen, Rubin, Neyman, Dantzig, Baer, & Siegelaub, 1964).

Diagnostic Agreement Studies

Although the literature says much about the importance of diagnosis in reading, little empirical data exists. The few diagnostic agreement studies in education suggest that groups of clinicians, working together, can produce mutually agreed-upon diagnostic statements (Lerner & Schuyler, 1973). Some studies in the medical literature, however, investigating the agreement of individual physicians' medical judgements, have revealed marked disagreement among physicians (Garland, 1959). Paton (1957) reported an error rate of 56% in diagnosing myocardial infarction based on autopsy results. In the diagnosis of pulmonary disease from x-ray photographs, the agreement of the average phsycian of the diagnosis was generally 80% with himself and 70% with other radiologists (Fletcher, 1952; Cochrane and Garland, 1952; Yerushalmy, 1955, 1969). Finally, in the diagnosis of various psychiatric disorders, there may be total disagreement among diagnosticians (Kendall, 1975).

The studies presented here derive from a program of empirical research on diagnositc problem solving in medicine (Elstein, Shulman, & Sprafka, 1978). These medical studies sought to capture the diagnostic methods used by highly skilled physicians who were presented with realistically simulated medical cases. The researchers concluded the physicians seemed to be hypothesis directed (generated successively more precise hypotheses of the patient's medical problems) and tested these hypotheses until a level of precision was reached that was satisfactory for treatment.

Understanding Diagnosticians' Problem-Solving Behavior: Study 1

The research to be reported in this paper was directed toward the understanding of diagnostic problem-solving behavior of expert practitioners in the field of education: reading specialists, learning disabilities specialists, and classroom teachers. The studies were based upon careful observation in a controlled setting.

The first study documented the characteristics of the interaction between reading specialists and a child with a reading problem to determine (1) what information these specialists collected, (2) what diagnostic categories they used, (3) what remedial actions they recommended, (4) how their diagnoses and remediations were related, and (5) how reliable these decisions were. The five subsequent studies were concerned with (1) replicating the original study, (2) examining the generalizability of the results of the initial study to other populations (three studies), and (3) examining the possible effects of artifacts of the data analysis procedures on the results (two studies).

The purpose of the *first observational study*, conducted in 1977, was to provide insight into the interaction between reading specialists and cases of reading difficulty. It was expected that the problem solving performance of these highly trained clinicians could serve as a model for the field of reading as it had in medicine. The experimental task for the individual clinicians was to diagnose and suggest remediation for simulated cases of reading difficulty.

Use of Simulated Cases

The use of simulated cases (as opposed to using a naturalistic setting with real children) insured that variation in clinician performance was attributable to variation in clinician, not case.

Research in medicine allays the concern that the diagnosis of simulated cases is a substantively different task from the diagnosis of real children. Norman and Tugwell (Note 3) support the assumption that important problem solving behaviors of clinicians can be elicited through simulated cases.

Each simulated case in this study consisted of collections of information about a child with reading problems. The simulated cases were based on real children in Grades 3-7 who had attended the Michigan State University Reading Clinic. They were considered by staff clinicians and outside consultants to be representative of reading problems commonly encountered in public schools. Across all the cases, the representative problems included sight word deficiencies, inadequate structural and phonetic analysis skills, inadequate oral reading fluency, and poor comprehension. Across all the cases, information about the child's achievement level, family and academic background, cognitive ability, reading ability, classroom behavior, and so on were presented in a variety of formats including test scores, completed test booklets, audio tapes, and written comments. Each simulated case was kept in a large file box and included an inventory of information (cues).

Four different simulated cases were created. Each simulated case had a replicate, a superficially disguised version prepared by making minor changes in the original case (Lee & Weinshank, Note 4). This made a total of eight different cases and allowed for a test/retest design.

Case 1: Stephen.

1. Initial contact information
   Age 8½ Grade 3 Taped interview
   Referred by teacher for reading problems

2. Potential for reading
   Good, at grade level or above
   Weschler Intelligence Scale for Children (WISC)
   Total I.Q.: 118 Verbal: 115 Performance: 118

3. Sight word vocabulary
   First Grade Dolch word list: 61%
   Slosson Oral Reading Test (SORT) placement: 2.2

4. Decoded word recognition
   Serious problem indicated--Gates-McKillop Subtest "Recognizing and Blending Common Word Parts" shows only 6 of 23 nonsense words were read correctly

5. Oral reading
   Inadequate fluency
   Durrell Analysis of Reading Difficulty, Oral Reading Subtest: Low second grade rate

6. Comprehension
   Listening comprehension above grade level
   Durrell Analysis of Reading Difficulty, Listening Comprehension Subtest: Grade 5; Durrell Silent Reading Subtest: third grade

Case 2: Donald.

1. Initial contact information
   Age 11 Grade 6 Taped interview
   Referred by teacher because of difficulties with reading related tasks

2. Potential for reading
   Adequate for grade level
   WISC Total I.Q.: 95 Verbal: 86 Performance 106
   Auditory acuity problem, audiological evaluation indicates significant hearing loss in upper frequency range

3. Sight word vocabulary
   Significantly below grade placement
   SORT placement: beginning 4th grade

4. Decoded word recognition
   Serious problem with decoding multisyllabic words
   Durrell Word Recognition and Analysis Subtest: both at 4th grade level

5. Oral reading
   Difficulty with phrasing
   Durell Oral Reading Subtest results indicate a word by word reader, rate at third grade level

6. Comprehension
   Problems in listening comprehension
   Durrell Listening Comprehension Subtest: Grade 4.5
   Problems in silent reading comprehension
   Gates-MacGinitie Comprehension Subtest: Grade level 2.9

Case 3: Mike.

1. Initial contact information
   Age 12 Grade 7 Taped interview
   Referred by parents who were concerned with his progress in areas involving reading, writing, and spelling.

2. Potential for reading
   Good for grade level and above
   WISC Total I.Q.: 105 Verbal: 111 Performance: 98

3. Sight word vocabulary
   Reasonably intact sight vocabulary for grade level
   Durrell Word Recognition: High 6th grade
   SORT grade equivalent: 6.8

4. Decoded word recognition
   Adequate to grade level
   Gates-McKillop Recognizing and Blending Common Word Parts: 20 of 23 nonsense words read correctly
   Durrell Word Analysis: mid sixth grade level
   Inadequate higher level decoding skills
   Gates-McKillop Syllabication Subtest: grade equivalent of 4.0

5. Oral reading
   Serious problems with fluency
   Durrell Oral Reading Subtest: rate is high fourth grade equivalent
   Gates-MacGinitie Speed & Accuracy Subtest: 5.2 grade equivalent

6. Comprehension
   Listening comprehension at grade level
   Durrell Listening Comprehension Subtest: 6th grade
   Silent reading comprehension below grade placement
   Gates-MacGinitie Comprehension Subtest: 5.5 grade equivalent

Case 4: Dan.

1. Initial contact information
   Age 9 Grade 4 Taped interview
   Referred by teacher and parents concerned about his basic reading skills and his lack of progress with reading related subjects.

2. Potential for reading
   Adequate for grade level reading and above
   WISC Total I.Q.: 101 Verbal: 103 Performance: 102

3. Sight word vocabulary
   Significantly below grade placement
   Dolch List: 71% on List 2

4. Decoded word recognition
   Severe problem with learning and application of decoding skills
   Durrell Word Analysis Subtest: Low first grade
   Durrell Visual Memory of Words Subtest: 1.5 grade equivalent

5. Oral reading
   Serious problem with rate
   Durrell Oral Reading Subtest: rate comparable to year-end second grader

6. Comprehension
   Listening comprehension above grade placement
   Durrell Listening Comprehension Subtest: 5th grade level
   Silent reading comprehension seriously depressed
   Iowa Test of Basic Skills Comprehension Subtest: grade 1.8 equivalency

The four case abstracts describe only part of the information available for each case. A complete listing of information for one of the cases is presented in Table 1.

Case replicates were prepared for all four simulated cases described above. The replicates were superficially disguised versions of Cases 1-4, prepared by making minor changes in each original case--changing names, using alternate forms of tests, re-recording tapes of oral reading, and so on.

The Study Participants

Participants were recuited from the most senior and most effective practicing clinicians in the mid-Michigan area. Recommendations were solicited from university faculty and/or school administrators. The candidates, a set of eight repeatedly recommended clinicians, were selected. All subjects had master's or doctoral degrees in reading and had been practicing as reading specialists for at least five years. They had received their training in various eastern and midwestern universities. All were paid at professional rates for their participation.

Table 1
Case 4: Dan
Cue Inventory

Information	Scores	Booklet	Audiotape	Directions
Physical				
Vision		X		
Audiometric record		X		
Background				
School record		X		
Teacher form		X		
School information		X		
Parent form		X		
Assessment				
Basic sight vocabulary (Dolch list)	X	X	X	X
Sentence completion		X	X	X
Reading Diagnostic tests (Gates-McKillop)				
-Recognition and blending common word parts	X	X	X	X
-Auditory blending	X	X		X
-Giving letter sounds	X	X	X	X
Auditory discrimination (Wepman)	X	X		X
Durrell listening/reading series, intermediate level				
-Vocabulary	X	X		X
-Paragraphs	X	X		X
Diagnostic analysis of reading difficulty (Durrell)				
-Oral	X	X	X	X
-Silent	X	X	X	X
-Listening comprehension	X	X		X
-Word recognition and word analysis	X	X	X	X
-Hearing sounds in words--primary	X	X		X
-Visual memory of words--primary	X	X		X
-Intermediate spelling--List 1	X	X		X
-Phonic spelling of words	X	X		X
Achievement test (Iowa Test of Basic Skills)				
-Vocabulary	X	X		X
-Reading	X	X		X
Graded word list (Slosson Oral Reading Test)	X	X		X
Reading achievement (Gates-MacGinitie)				
-Speed accuracy	X	X		X
Cognitive ability (Wechsler Intelligence Scale for Children)				
-Verbal	X			X
-Performance	X			X
-Full scale		X		

Design

Each clinician participated in three experimental sessions over a three-week period. Across the twenty-four sessions, each case/replicate was examined six times. Clinicians were randomly assigned to cases within the constraints of test/retest design and counterbalancing (See Table 2).

Table 2
Design of the Study

Clinician	Session		
	1	2	3
	Case		
A	1	4	1R
B	2	3	2R
C	3	1	3R
D	4	2	4R
E	4R	2	4
F	3R	1	3
G	2R	3	2
H	1R	4	1

R = replicate case

Procedures

Each session took place in a small room with a one-way mirror and consisted of an observation and a debriefing. Three people were present: the subject, an experimenter, and an observer who was also trained in reading. The experimental task for the subject was the diagnosis and remediation of a simulated case.

The observation. No time limit was imposed. The experimenter and subject sat near the one-way mirror, the observer sat on the other side of it. The experimenter began by helping the subject practice the experimental procedures using a simulated case different from the one to be used during the actual session. The session proceeded with the presentation of referral information and continued with the subject

requesting one piece of information at a time from the cue inventory. The experimenter would locate the information in the file box and present it to the subject. When the subject had collected as much information as desired, s/he was asked to write a diagnosis and an initial remedial plan. During consideration of the case, the subject was asked to verbalize his/her thinking, provided that doing so did not interfere with performance. The subject was encouraged to keep notes and proceed with his/her normal methods for diagnosing a case. Meanwhile, the observer on the other side of the mirror recorded on a standard observation form the information that was requested and the comments that were made by the subject.

The debriefing. The observer joined the subject and the experimenter. The three participants then reviewed the record of the subject's performance in the first part of the session. The observer reviewed with the subject each step of the interaction with the case, starting with the very first cue request and proceeding through the writing of the diagnosis. A set of three questions guided the debriefing for each cue: Why did you ask for this piece of information? What did it tell you? Did you have any hunches that were confirmed or ruled out, or was the information irrelevant? The observer was free to ask the subject to expand on any statement that the observer believed to be significant. The subjects' comments were recorded on a standard debriefing form. The intent was to reconstruct the clinician's thinking: Why were particular cues requested? How were specific cues interpreted? What hypotheses were generated by specific cues? Which cues confirmed or disconfirmed existing hypotheses?

Following the debriefing session, the subject had the opportunity to revise the written diagnosis in the event that the debriefing

session had altered his/her thinking about the case. Three products for the entire experimental session were (1) the standard observation form, which included cues collected, times of cue requests, and observer comments; (2) the standard debriefing form; and (3) the subject's written diagnosis and remediation including any additions made as a result of debriefing. For a more detailed description of the procedures for this study see Lee and Weinshank (Note 4).

Data Analysis

The clinicians' statements in the written diagnoses for each case were analyzed at two levels. First, the frequency of each diagnostic statement made across all sessions for a given case was tabulated (diagnostic commonality statistic). The proportion of sessions in which a statement was made provided an index of commonality for that statement. Second, the relationship between each pair of diagnoses was computed (diagnostic agreement statistics). The mean agreement statistics provided a measure of individual agreement for that case: interclinician agreement (between two clinicians on the same case) and intraclinician agreement (one clinician at two different times for the same case).

Diagnoses were compared in the following way. The natural language statements in each diagnosis were translated into a standard vocabulary (see examples in Table 3), established by project reading clinicians who sorted the diagnostic statements made in all sessions into equivalence classes. The more than two-thousand separate diagnostic statements made across all cases were grouped into 162 labeled classes. The interrater reliability was estimated by randomly reclassifying 10% of the 2,000 statements a second time: The result was a

Table 3
A Portion of the Standardized Diagnostic Categories

Number		Category[a]
3	Normal interests and behavior	O
36	At least average reading potential	S
39	Meaning vocabulary weak	W
50	Problem with visual memory	W
54	Reading not a meaningful act	W
60	Poor oral reading	W
64	No comprehension problem	S
65	Reading comprehension inadequate	W
71	Good use of context	S
72	Inconsistent use of context for word recognition	W
92	Sight words low	W
99	Insufficient visual discrimination and word scan	W
126	General statements about phonics	O
155	General statements about language	O
158	General home background statements	O

[a]W=weakness; S=strength; O=observation

Table 4
Conversion to a Standard Vocabulary of Three Diagnoses

Diagnoses	Clinician	Session	Simulated Case	Diagnostic Category[a]	Clinician Statement (Verbatim)
1.	2	1	4	99	He looks at the first letter or first few letters in the word.
				72	He guesses, many times wildly. The context of his guesses do not make sense.
				64	Scores indicate good comprehension.
				60	He doesn't read fluently.
				9	Has few automatic words.
2.	2	3	4R	99	Only looks at the first few letters in the word, ignoring the middle and the end.
				92	Storehouse of sight words low.
				71	Use of context is used well by Brian.
				50	Poor visual memory and sequential memory.
				155	He has a quantity of language but the quality may be somewhat lacking.
3.	6	2	4	92	Needs to increase sight vocabulary.
				71	One notable strength is his use of context.
				64	Able to extract meaning from the code.
				36	He has the potential to be an average or slightly above average reader.
				126	He has some initial consonants and blends.

[a]Refer to categories in Table 3.

75% placement of the statements in the identical categories the second time. The possibility of error in equating statements--that subjects might use different words to describe the same problem or that similar vocabulary might mask actual differences in meaning--was negligible. Our subsequent studies, in which vocabulary was controlled at the outset, showed this to be only a minor source of error (Hoffmeyer, Note 5; Stratoudakis, Note 6.)

The process of converting natural-language diagnostic statements into standardized diagnostic categories in illustrated in Table 4. The three diagnoses are all for the same case. The table presents sample diagnostic statements in both natural language and standardized categories. Thus, "He only looks at the first letter or first few letters in the word," was assigned to Diagnostic Category 99 (insufficient discrimination and word scan: weakness).

In order to determine commonality for each diagnostic statement for a case, a proportion was computed for each equivalence class in the standard vocabulary. The calculations of diagnostic commonality for three categories based on the sample diagnoses is shown in Table 5.

Table 5
Calculating Diagnostic Commonality[a]

Category		Diagnosis[c]		
		1	2	3
92	Sight words low (W)[b]	I	I	I
71	Good use of context (S)	A	I	I
60	Poor oral reading (W)	I	A	A

[a]The commonality statistic is calculated by dividing the number of times a category is included in a set of diagnoses by the total number of diagnoses in the set: C92=3/3=1.00, C71=2/3=.67, C60=1/3=.33.

[b]W=weakness; S=strength; O=observation.

[c]I indicates presence and A indicates absence from a diagnostic category.

As the table shows, the presence or absence of each diagnostic category is tabulated for each diagnosis. For example, Category 71, (good use of context) is absent from Diagnosis 1, but present in Diagnoses 2 and 3. The actual commonality statistic is calculated at the bottom of Table 4 (Footnote a). Note that these are only examples; the actual diagnoses contained many more statements. Further, each diagnostic commonality was calculated on the basis of six diagnoses, not three.

The commonality statistic gives no information about the extent of agreement between any two particular diagnoses for a given case. For this we used the agreement statistic. An agreement matrix would first list categories (by number of category) present in or absent from the diagnoses (numbers in upper part of each box) and the frequency. Below the matrix are the calculations for diagnostic agreement.

Table 6
Process for Determining Presence
or Absence of Diagnostic Agreement

		Diagnosis 1 Categories +	Diagnosis 1 Categories -	
Diagnosis 2	+	92, 99 N(+,+)=2	155, 71, 50 N(+,-)=53	A+B = 5
	-	72, 64, 60 N(-,+)=3	3, 36, 39, 54, 65, 126, 158 N(-,-)=7	C+D = 10
		A+C = 5	B+D = 10	

Note. +=present
-=absent
N=frequency of categories included in or absent from two diagnoses.

$$\text{Phi} = \frac{(A \times D) - (B \times C)}{\sqrt{(A+B)(C+D)(A+C)(B+D)}} \qquad \text{Porter} = \frac{A}{A+B+C}$$

$$\text{Phi}(1,2) = \frac{(2 \times 7) - (3 \times 3)}{\sqrt{(5 \times 10)(5 \times 10)}} \qquad P(1,2) = \frac{2}{2+3+3}$$

The frequencies are used to calculate Phi and Porter Coefficients.

The Porter Coefficient (bounded by 0 and 1) is easily interpreted. It is the number of diagnostic categories present in both diagnoses (A) divided by the number of categories present in either or both diagnosis (A+B+C). As the table shows, the Porter Coefficient for sample Diagnoses 1 and 2 is .25: Two out of eight diagnostic statements were agreed upon.

The Phi Coefficient is equivalent to the Pearson Product Moment Correlation when all scores are zero or one. Interpretation of the coefficient is usually similar to that of the Pearson: Zero indicates no relationship; one indicates perfect relationship. The baseline for interpreting the Phi in our work has been a study by Barrows, Feightner, Neufeld, & Norman who presented the same cases to 60 different physicians. The average Phi Coefficient for diagnoses of the same case was approximately .40. Since these diagnoses were based upon histories and physical exams with no verifying laboratory information available, they might be considered analogous to the type of information used by our subjects. Therefore, it could be argued that agreement of less than .40 would indicate a less than satisfactory state of affairs for the reading profession.

Methodological problems exist with the use of the Phi, since unequal marginal frequencies place bounds on the range of the statistic. Furthermore, one cannot assert that all entries in the "D" cell are the result of conscious decisions during both sessions to omit a diagnostic statement. One cannot clearly say whether the omission of a statement is the result of deciding to leave out that statement or never having considered it in the first place. Based on subsequent analysis of process, we have concluded that most of the entries in the "D" cell

represented diagnostic statements not considered by either clinician. Therefore, the "D" cell artifically inflates the correlation. The Porter Statistic avoids the problem of the inflated "D" cell by including only the statements actually made by one or both clinicians.

Agreement on information collected by each subject about the case was measured using the same procedures as the diagnoses. In addition, diagnostic processes such as hypothesis generation and time-related measures were examined.

Clinicians' Diagnoses

Each of the eight clinicians prepared a written diagnosis and remedial plan for three cases, yielding a total of 24 diagnoses and accompanying remediations. Following is a representative, complete diagnosis for Simulated Case Mike and a sample of the raw data from which we generated our results:

> Mike, a 12-year-old seventh grader with the capacity, family experiences, and background to perform at or above grade level in language-related subjects, scores substantially below level on standardized and objective-based tests.
>
> Several factors could have affected his ability to develop encoding and decoding skills: a speech problem that lingered into school years, farsightedness (in copying and reading from the board), partial auditory acuity problem, young kindergartener (sic).
>
> Mike's strengths are that listening skills appear to be close to his assigned grade (the test did not allow a ceiling for Mike's capacity). Mike uses his background and experience to make sense of reading as observed in passage-independent sentences of the Durrell Oral Reading Test. (i.e., Where does Henry go in the summer? --A camp. How far did water come in? --Pretty fast). Mike attempts to make sense from the book.
>
> Mike's weaknesses are inappropriate phrasing-fluency.
>
> He ignores punctuation.
>
> Reversals of letters occur both when he hears the sound and encoding and when he sees the symbol and decodes.

He deletes sounds when decoding, deletes symbols when encoding. Blends are identified correctly when he finds the blend symbol after hearing a word read.

Blends are incorrectly decoded--a vowel inserted or substitution of letters or one letter is ignored.

Related language skills are speech problems. Some speech problems with r and ending sounds were heard from interview.

Mike's handwriting was a combination of manuscript and cursive. The letters did not descend and the ascending letter barely ascended. The letters were poorly formed. Spacing was inappropriate.

Like all the others, this diagnosis is narrative in style and consists of three types of statements: strengths (child characteristics seen as helpful for reading), weaknesses (characteristics seen as being problematic for reading), and observations (statements that are either neutral or not clearly statements of strength or weakness).

Commonality

Little can be deduced from a single diagnosis. A better view of the common contents of the diagnoses is seen in the commonality statistics for the 162 standardized categories (Table 7). Table 7 includes only those categories mentioned in at least half the diagnoses for a given case. Only 16 categories met this minimum level of commonality. Categories most agreed upon were related primarily to reading potential, poor sight words, poor oral reading, poor word analysis skills, and poor attitude.

Mean diagnostic commonality provides an overall statistic representing the commonality of an entire study. It is obtained by averaging the commonality statistic across all cases and diagnostic categories. The maximum possible mean commonality is one (1). This mean value is obtained only when all diagnostic categories are

mentioned by all clinicians diagnosing the same case for all cases involved in the given study, that is, when all diagnoses for each case are identical. The minimum value of the mean commonality statistic is 1 divided by the number of diagnoses obtained for each case. For example, in the present study six diagnoses were obtained for each case. Hence, the minimum mean commonality value is 1 divided by 6, or .17. The minimum value is obtained when there is no agreement and all diagnoses are completely unique. The mean commonality obtained for this study is .26. Since this value is only .09 above the minimum, it can be seen that the extent of agreement among diagnoses is very low. Sixty percent of the standardized categories were mentioned only once. Fewer than 3% of the standardized categories were mentioned in half or more of the sessions.

Table 7

Diagnostic Categories Mentioned
Most Frequently Across All Four Cases

	Category	Case			
		1,1R	2,2R	3,3R	4,4R
92	Sight words low (W)	.33	.00	.83	.00
81	Phonics weak (W)	.33	.00	.00	.67
36	At least average reading potential (S)	.67	.33	.50	.67
60	Poor oral reading (W)	.50	.67	.33	.00
44	Adequate verbal skills (S)	.33	.50	.50	.00
86	Problem with vowels vowels (W)	.50	.33	.33	.00

Table 7 continued

Category	Case			
	1,1R	2,2R	3,3R	4,4R
21 Auditory acuity problem (w)	.00	.50	.00	.67
25 Attitude toward reading poor (W)	.50	.00	.00	.00
84 No problem with isolated letter sound skills (S)	.50	.00	.00	.00
11 Speech problem (W)	.00	.50	.00	.00
106 Problem with syllables(W)	.00	.50	.00	.00
115 Handwriting problem (W)	.00	.50	.00	.00
50 Problem with visual memory (W)	.00	.00	.50	.00
76 Poor word analysis (W)	.00	.00	.00	.50
109 Auditory discrimination problem (W)	.00	.00	.00	.50

Note. Statements listed are those mentioned in 50% or more of the diagnoses for a single simulated case.

Diagnostic Agreement

The analysis for diagnostic agreement showed that, on the average, two different clinicians agreed on only about 10% of the categories. When the diagnostic statements across two cases (case/replicate) for a single clinician were compared, on the average, about 20 percent of the categories mentioned by the clinician the first time s/he diagnosed a case were repeated when the replicate was diagnosed. The data indicate that written diagnoses across and within clinicians for the same case are extremely unreliable (Table 8).

Table 8

Reliability of Diagnoses and Cues
Mean Individual Agreement

	Written Diagnoses		Cues Requested	
	Interclinician	Intraclinician	Interclinician	Intraclinician
Phi				
Mean	-.10	.13	.13	.27
S.D.	.15	.10	.21	.34
Porter				
Mean	.10	.23	.30	.43
S.D.	.06	.07	.09	.13

Cue Collection Commonality

The major categories of cues most commonly collected were those that provided information about reading potential, oral reading, silent reading comprehension, listening comprehension, word recognition and word analysis, and home/school background information. Forty percent of the cues were collected once versus 60 percent of the diagnostic statements mentioned only once. Further, 30 percent of the cues were collected in half or more of the sessions (versus only 5 percent of the diagnostic statements being mentioned in half or more of the sessions).

Reading clinicians show a higher level of agreement on what data should be collected during the case work-up (cues) than they do in stating the diagnosis based upon such data (Table 8). On the average, any two clinicians agreed on 30 percent of the cues collected by both, in contrast to 10 percent agreement for diagnostic categories. On the average, the same clinician diagnosing the same case at two different times agrees on 43 percent of the cues collected in both sessions, in contrast to 23 percent for diagnostic statements.

The unexpectedly low diagnostic agreement in this study was startling, particularly since the clinicians who participated were highly trained (all but two had doctoral degrees) and had an average of ten years experience in their field. These findings raised three questions. First, was the low reliability a valid generalizable finding (i.e., not due to sampling error)? Second, would other professionals involved in the diagnosis and treatment of children with reading difficulties perform similarly? Third, did the statistics that indicated low agreement really reflect unreliable decision making, or was reliable performance being masked by artifacts of procedure or data analysis? Five further studies were carried out to address these questions.

Generalizability Studies

The five studies have been organized into two categories: those that focus on generalizing across populations and those that focus on modifying procedures and data analysis. This section focuses on the results of three studies that tested the generalizability to other populations of the finding of low diagnostic agreement for reading specialists. Each study description includes (1) the specific purposes of the study, (2) the population, (3) any design or procedural differences from the first observational study, and (4) results.

Learning Disability Study

A learning disability study was designed to (1) verify the initial observational study results with a different sample of reading specialists and new cases, (2) initially examine the interclinician agreement for learning disability specialists, and (3) compare the performance of these two groups of practitioners across cases in both fields (Van Roekel, Note 7).

Additional materials were prepared to accommodate the differing diagnostic training received by learning disabilities specialists. Since none of the existing cases contained the type of information necessary to complete a learning disabilities diagnosis, two simulated cases were developed for this study. The learning disability case was based on a real child with a learning disability; the reading case was one of the original reading cases modified to include learning disability measures that indicated no problem. Ten reading and ten learning disabilities clinicians each diagnosed both the reading and the learning disability cases. The subjects were randomly selected volunteers chosen from a list compiled by local school districts of qualified specialists. No intraclinician agreement measures were possible, since there was no test-retest design. The procedures for each session were identical to those in the 1977 observational study except that a 60-minute time limit on information collection was imposed for the first part of the session. The decision to limit the sessions to 60 minutes was based upon observations in the earlier study. First, most clinicians finished within one hour, and second, although some took longer than one hour, there was no correlation between time and diagnostic agreement. Subsequent studies

established that imposition of a time limit resulted in diagnostic agreements that were comparable to and sometimes higher than those in the original study.

The results of the learning disability study paralleled those of the 1977 study. Mean diagnostic commonality was .09 (S.D.=.04). In this study, the minimum commonality was .05 since 20 diagnoses were obtained for each case. The obtained value is only slightly higher than the minimum value. Despite lengthy individual diagnostic write-ups, few statements had a commonality of .5 or better (i.e., had been mentioned in at least half the diagnoses for a case). For the learning disability case, the only statements that were highly agreed upon were (1) weakness in gross/fine coordination and (2) problem with visual perception/discrimination/memory/motor skills. For the reading case, the most agreed upon statements were average intellectual potential, problem with attitude/interests, weak phonic analysis skills, and observations about contextual reading ability.

Interclinician agreement within each group revealed no differences. Both groups performed at a near zero level of reliability even within their own area of specialization. Only 5% of the diagnostic statements made about a case could be agreed upon by any two clinicians examining that case (Table 9).

Cue collection showed a higher level of agreement than did the diagnostic results. The mean commonality for cues collected was .17 (S.D.=.16); the mean interclinician Phi for cues was .20 (S.D.=.14).

The diagnostic and cue agreement of the reading clinicians in this study very nearly duplicated the intercorrelations reported for the 1977 study. The researchers therefore felt that sampling error

Table 9

Diagnostic Agreement of Reading (R) and Learning Disabilities (LD) Specialists

	R	LD	Both
Reading Case			
Mean Phi	.06	.04	.05
S.D.	.10	.12	.11
Learning Disability Case			
Mean Phi	.01	.07	.04
S.D.	.11	.13	.13

would be an unlikely explanation for the low diagnostic agreement of the reading specialists. Also, the learning disabilities clinicians were indistinguishable from the reading specialists in the low reliability of their diagnoses. It would appear that differences in the training of these two groups of professionals does not translate into differences in diagnostic performance.

Classroom Teachers

A classroom teacher's study was designed to (1) examine the inter-clinician agreement of classroom teachers, and (2) compare the performance of these teachers in experimental and classroom settings (Gil, 1979).

Moving from the study of highly trained reading and learning disabilities specialists to that of classroom teachers required the preparation of new simulated cases to accommodate a teacher's training and experience. Two cases were built around materials normally available to the classroom teacher: background information, samples of oral reading (tape recordings and accompanying transcriptions), and comprehension performance based on this reading. The cases did not include formal standardized measures of reading ability.

The design called for ten classroom teachers of diverse background (five from Michigan and five from Illinois) to (1) diagnose two different simulated cases and (2) discuss children in their own classrooms who matched the diagnostic profile of the simulated cases. Since the teachers did not diagnose the same case twice, no intraclinician analysis could be performed.

The procedures for each session were identical to those for the 1977 study with the exception of a 60-minute time limit imposed on data collection in the session's first part.

Diagnostic commonality for the classroom teacher study showed that only 6% of the total diagnostic statements were mentioned in three or more of the ten sessions. Mean commonality across diagnostic statements was .14 (S.D.=.09). The minimum mean commonality was .1 since a total of ten diagnoses were obtained for each case. The obtained value is only .04 above the minimum value. The diagnostic statements mentioned most frequently were poor comprehension, strength of major vocabulary concepts, sight words weak, ignores endings, sight vocabulary good, phonic skills weak, problems with oral reading, and lacks word attack skills.

The extent of individual (interclinician) agreement between two teachers on diagnostic statements was near zero (Table 10). Commonality for cues collected was slightly higher than for diagnostic judgments (mean=.19, S.D.=.28). Individual agreement on cues was near zero.

These findings paralleled those of the two preceding studies: Teachers, too, exhibited extremely low levels of agreement on the same case. It would seem that the training and experience of classroom teachers makes them no more reliable in the types of diagnostic decisions they must make than are the reading and learning disabilities

specialists.

The classroom component of this study was designed to assess whether teachers' diagnostic statements about real cases of reading disability in their classrooms were similar to their statements about simulated cases in the laboratory situation. Brief summaries of the simulated cases were presented to each teacher who was asked to identify a child in the classroom who most closely resembled one of the descriptions. Finally, the teacher was asked to describe the real child's reading problems. The analysis of classroom versus laboratory diagnoses showed that the diagnostic categories most frequently mentioned in the laboratory were also mentioned in the classroom. The data show that simulated cases in reading elicit the use of similar criteria as that found in natural settings.

Table 10

Mean Individual Diagnostic Agreement for Classroom Teachers

	Case	
	1	2
Mean Phi	.04	.03
Standard Deviation	.13	.11
Mean Porter	.05	.06
Standard Deviation	.06	.06

Group-Administered Simulated Cases

The purpose of the group-administered simulated case study was to test the generality of the findings of previous studies to a less restrictive information collection procedure (Stratoudakis, Note 6). This case contained the same categories of information as the individual one but with a smaller variety of measures. Case information was

presented in a looseleaf notebook. The critical procedural difference was that subjects, working alone but tested in a group, could interact with a case simultaneously under the supervision of just one experimenter, and they were free to thumb through the information in any order and at any rate.

The subjects in the study were 12 certified classroom teachers who had received a top grade in the graduate reading diagnosis course at Michigan State University. They examined three simulated cases in the individual or group format. Other variations on the procedures of the 1977 study included (1) a 30-minute limit on data collection, (2) subject translation of written diagnosis to standardized vocabulary, and (3) no debriefing session. Additional variations that applied only to the group-administered case were that (1) less information was available, (2) subjects examined case data without experimenter intervention, and (3) subjects examined the case in a group setting. All these variations on the 1977 procedure were designed to reduce the cost and complexity of data acquisition.

The commonality results show that once again, most diagnostic categories were mentioned by only one teacher. Mean commonality across diagnostic statements was .20 (S.D.=.13). The minimum commonlity was .17 since a total of six diagnoses were obtained for each case. The obtained value is only .03 above the minimum.

The categories mentioned most often differed little from those described in all previous studies. Categories mentioned most often included blends, sight-word recognition, word analysis, oral reading, fluency, and visual discrimination. The findings held for both the individual and group-administered simulated cases.

Individual diagnostic agreement remained largely unchanged despite the altered format and procedures, as shown in Table 11.

Table 11

Mean Individual Diagnostic Agreement
for Group Administered Simulated Cases

	Two Simulated Cases (Experimenter Controlled)	Two Simulated Cases (Experimenter Controlled and Subject Controlled)
	(N=39)	(N=30)
Phi		
Mean	.07	.10
S.D.	.06	.06

Mean commonality on cues was .34 (S.D.=.20); mean individual agreement was .17 (S.D.=.16). In general, these last three studies seem to indicate that (1) the low-agreement findings were not uniquely characteristic of the particular clinicians in the 1977 study; (2) the low agreement findings do seem to characterize the decision making of other education professionals; and (3) the low agreement is probably not an artifact of the particular cases and procedures used in the 1977 study.

Replication Studies

In this section we will present the results of two operational replicates (Borg and Gall, 1979) of the 1977 study. These replicate studies were aimed at examining the artifacts of experimental procedure and data analysis that could have accounted for the finding of low reliability in the earlier studies. The description for each study will include (1) the purposes of the study, (2) the population, (3) the design and procedures only as they differed from the first observational study, and (4) results.

Study of Vocabulary Standardization Procedures

The purposes of the vocabulary study were to (1) provide an operational replication of the 1977 study and (2) investigate methodological problems in experimenter translation of natural language diagnostic statements into a standard vocabulary (Hoffmeyer, Note 5). A concern was that in categorizing the clinicians' natural language statements, the experimenters might have failed to see equivalences. In that case, statements that were actually describing the same thing would be coded as being dissimilar, and agreement would appear to be very low. Conversely, statements actually describing different problems might incorrectly be equated.

The design and simulated cases of the study were identical to that of the 1977 study. The eight subjects were senior clinicians who have been nominated by university faculty; all but two had doctoral work in reading. Each subject diagnosed three simulated cases, the first and third being the same case with minor changes (i.e., name, sex, etc.) Some procedures differed from the other studies.

First, participants were given a 45-minute time limit for data collection to reduce the possibility that subjects might become confused by an information overload. Second, subjects were not provided with an inventory of available information to address the possibility that the inventory was stimulating subjects to ask for information that they would not otherwise request. Third, subjects translated all natural language statements in their written diagnoses to a standardized checklist. The checklist was empirically derived from clinician statements in the preceding observational studies. Fourth, the debriefing session was conducted via a written questionnaire.

Once again, commonality across cases focused on the same diag-

nostic categories as in the original investigations: (1) average intellectual potential, (2) poor oral reading, (3) sight words low, (4) phonics weak, (5) poor word analysis skills, and (6) problem with auditory acuity. There was some agreement on two additional categories: problem with comprehension and poor attitude toward reading. For any given case only a few statements could be gleaned that represented commonality on case characteristics. Mean commonality on diagnostic statements was .24 (S.D.=.13). The minimum mean commonality is .17 since a total of six diagnoses were obtained for each case. The obtained value is only .07 above the minimum value.

The results for interclinician agreement showed a slight increase over the original study (Table 12).

Table 12

Mean Individual Diagnostic Agreement
for Study of Vocabulary Standardization Procedure

	Diagnostic Agreement	
	Interclinician	Intraclinician
Phi		
Mean	.11	.32
S.D.	.08	.11
Porter		
Mean	.08	.21
S.D.	.05	.07

The results for intraclinician agreement remained essentially the same. Therefore, while the original method of standardizing vocabulary may have produced slight underestimation of the individual agreement results, the differences are clearly not great enough to implicate the translation procedure as the explanation for the generally low reliability.

Again there was more commonality on *which* cues to examine than on *what* diagnostic statements to derive from them. Mean commonality on information requested was .43 (S.D.=.26). Individual agreement was lower: interclinician mean was .33 (S.D.=.15); intraclinician mean was .42 (S.D.=.20).

Study of Diagnosis and Remediation

This study had several purposes (Weinshank, 1982). First, it provided an operational replication of the 1977 study. Second, it studied the reliability of remediated diagnoses. In all previous studies, the analysis was performed on all diagnostic statements without reference to the remediation. In this study, the analysis was altered to additionally examine the reliability of those diagnostic categories for which remedial recommendations were made. The original method might have inadvertently "swamped" substantial agreement by combining a few reliable, important diagnostic categories linked to remediation with a larger number of unreliable, unimportant ones that would not be linked with remediation. Third, the study examined the reliability of remedial statements themselves. Since remediation prescriptions lead to actions, it was conjectured that there would be greater reliability with respect to remediation prescriptions chosen. Finally, the relationship between diagnosis and remedial plans was examined.

The design and simulated case materials of the study were essentially identical to the 1977 study. The subjects were practicing reading teachers, all of whom held master's degrees and had an average of 11 years of experience each. Four had received their graduate training in Michigan, the other four in Illinois.

The major procedural differences between this study and the original one were (1) the use of a standardized diagnostic checklist for subject translation of diagnostic statements, (2) the use of a standardized remedial checklist for subject translation of remedial statements, (3) development of new debriefing procedures used only in the final session and focused on reasons for associating or not associating diagnostic and remedial statements, (4) the addition of analysis of the reliability of diagnostic statements linked to remedial plans, and associations between specific diagnostic and remedial statements.

This study examined five products for each session. Two of these were diagnostic and cue performance records examined in all previous study. Further analysis was performed on (1) remedial statements, (2) diagnostic statements linked to remediation prescription, and (3) the associations made between specific diagnoses and specific remediation prescriptions (diagnostic/remedial associations).

The mean commonality for four of these products is summarized in Table 13.

The minimum mean commonality is .17 since six diagnoses were obtained for each case. As in the previous studies, there is very little differences between the obtained and the minimum mean commonality.

Ten percent of the diagnostic categories that were mentioned accounted for whatever commonality existed across cases. Those categories were (1) at least average intellectual potential;

problems with (2) word recognition, (3) word analysis, (4) oral reading, (5) silent reading, (6) comprehension, (7) auditory/visual acuity, (8) auditory discrimination, and (9) affect.

Table 13
Mean Commonality: Study of Diagnosis and Remediation

Products	Mean	S.D.
Diagnosis	.26	.15
Remediation	.24	.13
Cues	.43	.23
Remediated Diagnosis	.24	.13

Similarly, ten percent of the remedial categories mentioned accounted for whatever commonality existed across cases. Those categories were (1) sight words, (2) phonetic analysis, (3) structural analysis, (4) oral reading, (5) visual problems, (6) comprehension, and (7) motivation.

Table 14 shows the results for interclinician and intraclinician agreement. Overall, global diagnostic reliability and cue reliability remained similar to the other studies; commonality on remedial actions to be used was also unreliable but individual agreement on remedial actions showed slightly more reliability; agreement on precisely which diagnoses warranted treatment was no better than for the global diagnosis, and the relationship between specific diagnoses and specific remediations was shown to be near zero.

In general, replication studies seem to show (1) that the finding of low reliability can be replicated with other samples from a population similar to that of the first study, (2) that the low reliability was not induced by experimenter error during the standardization of

the natural language diagnoses, and (3) that remediations and the diagnoses linked to remediations were no more reliable than global diagnoses.

Table 14

Mean Individual Diagnostic Agreement:
Study of Diagnosis and Remediation

	Clinician Agreement			
	Inter		Intra	
	Phi	Porter	Phi	Porter
Diagnosis	.16	.11	.23	.14
Remediation	.14	.10	.29	.20
Remediated Diagnosis	.13	.08	.22	.14
Remedial/Diagnostic Associations	.18	.00	-.10	.02
Cues	.24	.29	.31	.49

Measures of Diagnostic Process

In addition to the results presented for the preceding six studies, some of the processes by which the subjects reached their decisions have been documented. Because of variations in procedures, process statistics are available only for the original 1977 study, the learning disability study, and the classroom teacher study.

First, the data show a broad range of time taken for cue collection and examination. About one-fourth of the subjects took 30 minutes or less, while another fourth took an hour or more.

Second, the number of cues requested varied widely. In the three studies (N=84), the subjects requested a mean of 33 items from their case inventory (range=11,89). The number of cues collected was only moderately related to the length of time the subject took to examine them (r=.37). A subject who took an hour or more to examine requested

information did not necessarily collect many more cues than a subject who took half an hour or less. In sum, there appears to be no significant correlation between time taken, number of cues collected, and diagnostic agreement.

Subjects requested cues in order to formulate a diagnosis. During this process they were asked to verbalize their thinking. For the 1977 study, statements of hunches or hypotheses were extracted from these verbalizations. Hypotheses were analyzed with respect to when they were initially considered: during the first, second, third or fourth quarter of the session. The results show that almost half of all hypotheses were generated in the first quarter. Hypothesis generation declined dramatically from this point to less than 10 percent in the final quarter. However, this was not the pattern for cue collection across quarters. The number of cues requested remained essentially constant across all four quarters.

The mean results for cue and hypothesis performance across cases in the three studies are summarized in Figure 1.

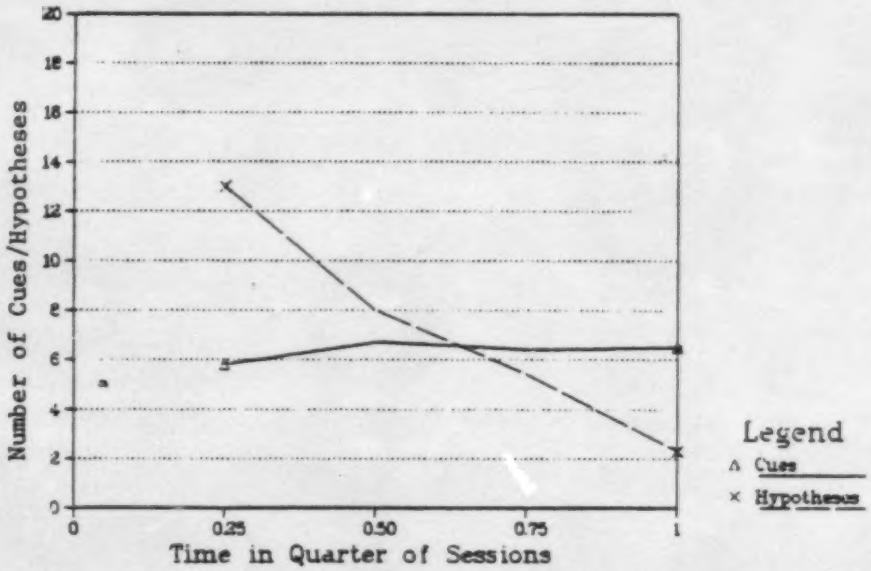


Figure 1. Mean results for cue and hypothesis performance across cases in three studies.

The figure shows that the clinicians continued to collect information long after most hypotheses had been generated. It may be that the subjects needed additional information in order to confirm or reject existing hypotheses. Alternatively, they may have continued to collect data to increase their confidence in judgments already made.

The process of cue collection, hypothesis generation, and verification culminated in the subjects' writing of diagnoses to contain some subset of the hypotheses considered. In the 1977 study, an average of 39 percent of the hypotheses considered were confirmed and stated in the diagnosis as characterizing the case. That is, of an average of 31 hypotheses, 12 were carried over into the written diagnosis. The remaining hypotheses are assumed to have been rejected or simply forgotten.

As a result of our analyses of process, our general impression is that the professionals studied did not have an overall strategy or framework for reliably linking cues with hypotheses and hypotheses with diagnoses.

Summary of Findings

There are two related findings across all studies: commonality and individual agreement are both very low. The findings on commonality show that most statements in a written diagnosis and remediation for the same case are mentioned by only one clinician. Mean commonality for all studies is summarized in Table 15. By far the most frequently mentioned categories within and across studies are potential for reading, sight words, word analysis, oral reading, attitude, comprehension, visual discrimination, and auditory acuity.

Table 15

Mean Diagnostic Commonality Across Studies

Study	Mean	S.D.
Initial Observational	.26	.14
Learning Disabilities	.08	.04
Classroom Teachers	.14	.09
Vocabulary Standardization	.24	.13
Group Admin. Cases	.21	.13
Diagnosis & Remediation	.26	.15
Grand Mean	.20	.11

A second major finding across all studies is the low individual diagnostic agreement. Individual agreement data for all studies is summarized in Table 16.

Table 16

Mean Individual Diagnostic Agreement Across Studies

Study	Clinician			
	Inter Phi	Inter Porter	Intra Phi	Intra Porter
Initial Observation	-.10	.10	.13	.23
Learn Disabilities	-.02	.03		
Classroom Teachers	-.04	.06		
Vocab. Standard	.11	.08	.32	.21
Group Administered Cases	.09		.14	
Diagnosis & Remediation	.15	.11	.23	.16
Grand Mean	.03	.08	.21	.20

Low diagnostic and remedial reliability appears to be a robust phenomenon. While the reliability for two clinicians diagnosing the same case is lower than that for a single clinician on a case and its replicate, both levels of reliability are low. The mean across all studies is not much better than chance.

Additional analyses performed on cue records show that the reliability of choosing information is consistently greater than that for diagnosis and remediation, both for commonality (mean=.32) and individual agreement (mean Phi=.18, mean Porter=.23).

Analysis of the diagnostic and remedial process in three studies revealed wide variability in clinician performance. The total time taken for cue collection and the number of cues that were collected varied greatly among participants and did not significantly correlate with diagnostic agreement. Two behaviors that did appear to be constant across clinicians were (1) hypothesis generation decreased sharply across each session but (2) the number of cues collected remained constant across the entire session.

Discussion

Reading experts show near universal agreement on the importance of diagnosis as the basis for the remediation of reading problems. Authorities are also in broad agreement on the following major contents of a diagnosis though weighting them differently: (1) determination of overall potential for reading, (2) performance on specific skills, and (3) exploration of causal factors. The literature shows similar agreement on the conduct of the diagnosis. Depending on the particular case, diagnosis may be conducted (1) by the classroom teacher using group tests, (2) by the reading specialist using individualized reading diagnosis instruments in the school setting, and (3) by a group of professionals from diverse fields in a clinic setting.

The empirical studies reported in this paper show that reading professionals, as a group, produce aggregate diagnoses that include statements about reading potential, strengths and weaknesses in skills, and selected causal factors (hearing, vision, and attitude), thus conforming to the recommendations of authorities in the field.

Individually, the diagnoses show significant deviations from the recommendations of the experts. First, the diagnoses include a large number of one-time-only statements of questionable relevance to remediation. Second, the diagnoses fail systematically to mention the reading skills of greatest import to remediation. Third, even when important skills are mentioned in the diagnosis, these statements are not reliably linked with treatment.

One possible explanation for these low agreements might be found in the use of simulated cases in an experimental environment. However, the use of real children in a natural setting introduces factors that might further decrease agreement since a child's behavior and performance would be expected to change, thereby introducing unreliability in the data base for a case.

The differential effects of using real and simulated cases has been studied in medicine. No differences were found when diagnoses are compared for (1) people with real medical problems, and (2) people coached to simulate the same medical problems (human simulation). Further, in studies comparing human simulation of medical problems with simulated cases similar in format to those used in our studies, differences were found in procedure but not in the final diagnoses.

A second possible explanation for the low diagnostic agreement found in our studies lies in the nature of the training that reading specialists receive. A comparison of programs in medicine and reading is instructive here. Medical training is based upon (1) an organized body of empirically based knowledge that relates specific remedies to specific problems; (2) systematic techniques governing the collection of cues; and (3) perhaps most importantly, the supervised diagnosis, treatment, and follow up of thousands of cases. By contrast, training in reading tends to lack all three of these above characteristics, and instead,

has (1) non-empirically verified theoretical concepts, (2) idiosyncratic cue collection techniques, and (3) supervised diagnosis, remediation, and follow-up of few cases.

The importance of diagnostic reliability in establishing the connections between prediction and outcomes requires that action be taken to provide the kind of training that will support teacher learning in this skill. Results of training studies (Sherman, Weinshank & Brown, Note 8; Gil, Polin, Vinsonhaler & VanRoekel, Note 9; Polin, Note 10) indicate that inter-clinician reliability on key diagnostic categories (instant word recognition, decoded word recognition, word meanings, oral reading, silent reading comprehension, listening comprehension, and attention/motivation) can be increased substantially. Instruction emphasizing external decision aids (including computer support), an explicit model of the diagnostic process, and practice with feedback on a variety of simulated and real cases appears to provide a powerful common heuristic for data collection and interpretation.

Perhaps the major implication of the studies concerns the question of the place of diagnosis itself in the correction of reading problems. In the first place, diagnosis as presently conducted should not be continued. Should diagnosis be discarded, then, as a precursor to remediation? The answer depends upon whether differentially effective remediations exist, (i.e., remediations that are more effective for one problem than another). If remediations are uniformly effective, there is clearly no need for a diagnosis to guide the selection of treatment. However, if differential treatments do exist, then reliable diagnoses are indispensible.

As noted in our introduction, differentially effective remediations are empirically discernible only when reliable diagnoses are used

in the evaluation of their effectiveness. Given reliable diagnosis, the stage is set for the aggregation of the data necessary to establish differentially effective remediation across children and problems.

Finally, if the results reported in this paper prove the rule, we must not castigate reading clinicians; similar results of low agreement have been encountered in medicine and psychology. Instead, we must seek a better understanding of the causes of low diagnostic agreement and methods by which we may use this knowledge to improve the training and decision making of reading specialists, and ultimately better practice.

Reference Notes

1. Austin, M.C. Diagnostic teaching: The primary grades. In J.H. Root (Ed.), Diagnostic teaching: Methods and materials. Syracuse, New York: Syracuse University, School of Education, 1965.

2. Bateman, B. The role of individual diagnosis in remedial planning for reading disorders. In E.O. Calkins (Ed.), Reading Forum (NINDS Monograph No. 11). Bethesda, Maryland: U.S. Department of Health, Education, & Welfare, 1971. (NIH 73-44)

3. Norman, G.R., & Tugwell, P. The validity of simulated patients. Presentation at the 1981 Research in Medical Education conference, Washington, D.C., 1981.

4. Lee, A., & Weinshank, A. Case production and analysis: CLIPIR pilot observational study of reading diagnosticians (Research Series No. 14). East Lansing, Michigan: Institute for Research on Teaching Michigan State University, 1978.

5. Hoffmeyer, L. The process and outcomes of diagnostic problem solving among reading clinicians. Unpublished doctoral dissertation, College of Education, Michigan State University, 1980.

6. Stratoudakis, J. The construction and use of paper problems to observe the diagnostic problem solving behavior of reading clinicians. Unpublished doctoral dissertation, College of Education, Michigan State University, 1980.

7. Van Roekel, J. The problem-solving behavior of reading specialists and learning disabilities resource teachers. Doctoral dissertation in progress, College of Education, Michigan State University, 1981.

8. Sherman, G.B., Weinshank, A., & Brown, S. Training reading specialists in diagnosis (Research Series No. 31). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1979.

9. Gil, D., Polin, R., Vinsonhaler, J., & Van Roekel, J. The impact of training on diagnostic consistency (Research Series No. 67). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1980.

10. Polin, R. A study of preceptor training of classroom teachers in reading diagnosis (Research Series No. 110). East Lansing, Michigan: Institute for Research on Teaching, Michigan State University, 1981.

References

Bond, G.L. Diagnostic teaching in the classroom. In D.L. DeBoer (Ed.), Reading diagnosis and evaluation. Proceedings of the 13th annual convention of the International Reading Association (Vol. 13). Newark, New Jersey: International Reading Association, 1970.

Bond, G.L., & Tinker, M.A. Reading difficulties, their diagnosis and correction (2nd ed.). New York: Appleton-Century-Crofts, 1967.

Borg, W.R., & Gall, M.D. Educational Research. New York: Longman, 1979.

Carter, H.L. The teacher's diagnosis of the disabled reader. In W.K. Durr (Ed.), Reading difficulties: Diagnosis, correction and remediation. Newark, Delaware: International Reading Association, 1970.

Carter, H.L., & McGinnis, D.J. Diagnosis and treatment of the disabled reader. Toronto: MacMillan Co., 1970, Chapters 3 & 9.

Cochrane, A.L., & Garland, L.H. Observer error in the interpretation of chest films: International investigation. Lancet, 1952, 2, 505-509.

Collen, M., Rubin, L., Neyman, J., Dantzig, G., Baer, R., & Siegelaub, A. Automated multiphasic screening and diagnosis. American Journal of Public Health, 1964, 54(5), 741-750.

Dauzat, S.W. Informal diagnosis: The nucleus. In R.A. Earle (Ed.), Classroom practice in reading. Newark, Delaware: International Reading Association, 1977.

Dietrich, D.M. A diagnostic approach to corrective reading in the classroom. Paper presented at the annual meeting of the International Reading Association, Detroit, May 1972. (ERIC Document Reproduction Service No. 063 576)

Ekwall, E. Diagnosis and remediation of the disabled reader. Boston, Massachusetts: Allyn & Bacon, 1976.

Elstein, A., Shulman, L.S., & Sprafka, S. Medical problem solving: An analysis of clinical reasoning. Cambridge: Harvard University Press, 1978.

Farr, R. Reading diagnosis: Trends and issues. Paper presented at the meeting of the Nassau Reading Council, Uniondale, New York, March 1971. (ERIC Document Reproduction Service No. 050 899)

Fletcher, C.M. Clinical diagnosis of pulmonary emphysema: Experimental study. Proceedings of the Royal Society Medicine, 1952, 45, 577-584.

Garland, L.H. Studies on the accuracy of diagnostic procedures. American Journal of Roentgenology, 1959, 82, 25-38.

Gil, D. The clinical problem solving behavior of classroom teachers as they diagnose a child's reading performance in laboratory and classroom. Unpublished doctoral dissertation, College of Education, Michigan State University, 1979.

Guszak, F.J. Diagnostic reading instruction in the elementary school. New York: Harper & Row, 1972.

Harris, A.J. Ten years of progress in remedial reading. Journal of Reading, 1977, 21(1), 29-35.

Harris, A.J. The diagnosis of reading disabilities. In A.J. Harris & E.R. Sipay (Eds.), Readings on reading instruction (2nd ed.). New York: David McKay Co., Inc., 1972.

Karlsen, B. Reading: Assessment and diagnosis of abilities. In P. Lamb & R. Arnold (Eds.), Reading: Foundations and instructional strategies. Belmont, California: Wadsworth, 1976.

Kendall, R.E. The role of diagnosis in psychiatry. Oxford, Great Britain: Blackwell Scientific, 1975.

Kennedy, E.C. Classroom approaches to remedial reading. Itasca, Illinois: Peacock, 1971.

Lerner, J., & Schuyler, J. Computer applications in the field of learning disabilities. Final Report, U.S. Dept. of Health, Education & Welfare, August 1973.

Monroe, M. General principles of diagnosis of reading disabilities. In D.G. Schubert & T.L. Targerson (Eds.), Readings in reading: Practice theory research. New York: Crowell, 1968.

Natchez, G. (Ed.), Children with reading problems. New York: Basic Books, 1968.

Newman, H.(Ed.), Reading disabilities: Selections on identification and treatment. New York: Odyssey, 1969, Chap. 2.

Olson, J.P., & Dillner, M.H. Learning to teach reading in the elementary school utilizing a competency based instructional system. New York: MacMillan, 1976.

Otto, W., McMenemy, R.A., & Smith, R.J. Corrective and remedial teaching (2nd ed.). Boston: Houghton Mifflin, 1973.

Paton, B.C. The accuracy of diagnosis of myocardial infarction. American Journal of Medicine, 1957, 23, 761-768.

Peters, C.W. Diagnosis of reading problems. In W. Otto, N.A. Peters, & C.W. Peters (Eds.), Reading problems: A multidisciplinary perspective. Cambridge, Massachusetts: Addison-Wesley, 1977.

Rabinovitch, R. Differential diagnosis in children with reading retardation. In J.H. Root (Ed.), Diagnostic teaching: Methods and materials. Syracuse, NY: Syracuse University, School of Education, 1965.

Sawyer, R. Diagnosis for the classroom teacher. Paper presented at the conference of the College Reading Association, Knoxville, Tennessee, April 1968. (ERIC Document Reproduction Service No. 029 754)

Sheldon, W.D. Specific principles essential to classroom diagnosis. In D.G. Schubert & T.L. Targeson (Eds.), Readings in reading: Practice theory research. New York: Crowell, 1968.

Smith, C.B. Correcting reading problems in the classroom. Newark, Delaware: International Reading Association, 1969.

Smith, C.B., Carter, B., & Dapper, G. Treating reading difficulties: The role of the principal, teacher, specialist, administrator. Washington, D.C.: U.S. Department of Health, Education and Welfare, 1970.

Spache, G.D. Diagnosing and correcting reading disabilities. Boston: Allyn & Bacon, 1976.

Spache, G.D., & Spache, E.B. Reading in the elementary school (3rd ed.). Boston: Allyn & Bacon, 1973.

Strang, R. Diagnostic teaching of reading. New York: McGraw-Hill, 1964.

Strang, R. Levels of reading diagnosis. Educational Forum, 1969, 33(2), 187-191.

Swalm, J.E. Diagnostic teaching: A method for assessing reading skills. In M. Kling (Ed.), Language development for the classroom and remedial reading. Proceedings of the Fall 1972 Rutgers University Reading Conference, New Brunswick, 1973. (ERIC Document Reproduction Service No. 140 260)

Weinshank, A. The reliability of diapnostic and remedial decisions of reading specialists. Journal of Reading Behavior, 1982, 14(1), 33-50.

Wilson, R.M. Diagnostic and remedial reading for classroom and clinic (3rd ed.). Columbus, Ohio: Merrill, 1977.

Yerushalmy, J. Reliability of chest radiography in the diagnosis of pulmonary lesions. American Journal of Medicine, 1955, 89, 231-240.

Yerushalmy, J. The statistical assessment of the variability in observer perception and description of roentgenographic pulmonary shadows. Radiologic Clinics of North America, 1969, 7, 381-391.

APPENDIX

Standardized Diagnostic Categories

1. Nervous, frustrated child
2. Work appears organized
3. Normal interests and behavior
4. Aggressive, impulsive child
5. Attending behavior needs improvement
6. No attention span problem
7. Poor self concept
8. No risk-taking behavior
9. Not afraid to try
10. No speech problem
11. Speech problem
12. Speech and hearing problems related
13. Suspected learning disability
14. Health problems in school
15. Small physical size
16. No physical problems
17. Visual acuity and farsightedness problem
18. No vision problem
19. Possible vision problem
20. No auditory acuity problems
21. Auditory acuity problem
22. Reading problem may be physical
23. Dependent reader
24. Lack of motivation for reading
25. Attitude toward reading poor
26. Normal motor skills
27. Visual motor skills adequate
28. Visual/motor problems
29. Normal home environment
30. Parent-school cooperation inadequate
31. Family values reading
32. Parents are cooperative
33. Parents don't like to read
34. Parent anxiety about school and reading
35. Parents' educational background
36. At least average reading potential
37. Good potential for learning sight words
38. Low average reading potential
39. Meaning vocabulary weak
40. Not working up to potential
41. Comparative statements on verbal and performance scores on WISC
42. Poor listening comprehension
43. Problem with verbal skills
44. Adequate verbal skills
45. Vocabulary adequate
46. Problem with auditory memory
47. Auditory memory adequate
48. Is problem with hearing or with reproduction
49. Visual memory good
50. Problem with visual memory
51. Reading problem

52. Not bad in reading
53. Not performing at grade level
54. Reading not meaningful act
55. Not accurate
56. Operating at first grade level
57. Word by word reader
58. Reading rate slow
59. Sight reader
60. Poor oral reading
61. Silent reading score equal to oral reading
62. Second grade oral reading score
63. Oral reading makes him look better
64. No comprehension problem
65. Reading comprehension inadequate
66. Good listening comprehension/poor reading comprehension
67. Second grade reading comprehension skills
68. Comprehension higher than expected
69. Reading vocabulary higher than comprehension
70. Poor word identification impedes comprehension
71. Good use of context
72. Inconsistent use of context for word recognition
73. Limited use of contextual analysis
74. Most errors surface errors
75. Did not make deep structure errors
76. Poor word analysis skills
77. Guesses at unknown words
78. No risk taking in word analysis
79. No independent analysis
80. Uses word recognition skills when cued
81. Phonics weak
82. Poor letter sound association
83. Some phonetic ability
84. No problem with isolated letter sound skills
85. Isolated (phonics) instruction problem
86. Problem with vowels
87. Knows short vowels in isolation
88. Consonant blends not a problem
89. Problem with blends
90. Adequate word recognition and analysis
91. Does not use ending cues
92. Sight words low
93. Sight vocabulary not automatic
94. Visual discrimination not a problem
95. Reversal problems
96. Configuration application for word recognition inappropriate
97. Fourth grade sight vocabulary
98. Poor discrimination of redundant letter combinations
99. Insufficient visual discrimination and word scan
100. Ignores or confuses middle letters
101. Inconsistent use of beginning and ending letters and sounds
102. Uses beginning letters and sounds for word identification
103. Problem with discrimination of visually similar words
104. Affixes no problem
105. Inadequate structural analysis skills
106. Problem with syllables
107. Understands idea of syllables

108. Auditorially discriminates sounds in words
109. Auditory discrimination problem
110. Problem with auditory stream
111. Problem synthesizing parts into wholes
112. Can blend short words
113. Difficulty blending sounds into words
114. Spelling/writing problems
115. Handwriting a problem
116. Writing problem
117. Spelling problems
118. SAT scores show growth
119. Needs structured instruction
120. Inconsistent instruction
121. Inappropriate instructional materials
122. Late identification of reading problem
123. Overplacement in school reading materials is a problem
124. Math near grade level
125. Problem with numbers
126. GSA phonics
127. GSA word chunking
128. GSA structural analysis
129. GSA configuration
130. GSA reversals and deletions
131. GSA blending
132. GSA blends
133. GSA speed and accuracy
134. GSA sight words
135. GSA word recognition and word analysis
136. GSA words in isolation
137. GSA discrimination and memory
138. GSA visual memory
139. GSA performance in reading compared to potential
140. GSA deep structure
141. GSA potential
142. GSA verbal and non-verbal performance on WISC
143. GSA affect
144. GSA hearing
145. GSA vision
146. GSA speech
147. GSA perception
148. GSA physical problems
149. GSA health
150. GSA motor performance
151. GSA home background
152. GSA teacher relationship
153. GSA reading as a school subject problem
154. GSA grade level of performance
155. GSA language
156. GSA comprehension
157. GSA math
158. GSA handwriting
159. GSA spelling
160. GSA retrieval
161. Non-specific observations and questions
162. Unique statements